Conference contribution :
Islam, M., Chaussalet, T., Ozkan, N. & Demir, E. (2010). *An approach to exploring the effect of weather variations on chronic disease incidence rate and potential changes in future health systems.*(pp. 190-196).

http://dx.doi.org/10.1109/CBMS.2010.6042639

An Approach to Exploring the Effect of Weather Variations on Chronic Disease Incidence Rate and Potential Changes in Future Health Systems

Muhammad Saiful Islam[1], Thierry Chaussalet[1], Nazmiye Ozkan[2], Eren Demir[1]
[1] *Health and Social Care Modelling Group,* [2] *Policy Studies Institute, University of Westminster, London, UK.*
muhammad.islam@my.westminster.ac.uk, {chausst, E.Demir1}@westminster.ac.uk, Ozkann@psi.org.uk

Abstract

The ecology and the environment of the world are changing due to the changing patterns of the meteorological factors. Scientific recognition confirms that such variation have already started affecting health and disease prevalence. In England, circulatory system and respiratory diseases are known to have high incidence and mortality rate, especially during the winter. This paper highlights the changing trends of incidence rate of chronic obstructive pulmonary disease (COPD) and some meteorological variables (monthly temperature and total rain) for a selected region (England North) for1997 to 2004.Zero-inflated Poisson regression model was used to observe whether meteorological variables are significantly related to COPD admissions counts. This can be useful to better understand the factors associated with prevalence rates for specific disease categories. Our preliminary findings support the case for developing advanced methodologies to examine regional disparities in our healthcare system due to weather changes.

1. Introduction

The last decade was the warmest on record, followed by the 1990s and then 1980s. The five warmest years were, in ascending order, 2002, 2003, 2005, 1998 and 2009 [1]. Currently there is a growing scientific consensus [2] that the change in meteorological variables is already adversely affecting health, and these effects will be unevenly distributed throughout the world, making children and older people more vulnerable than ever [3].

One unit of increase in temperature may result in a 1-4% increase in mortality in European countries [3]. Older people are most at risk from heat-stroke and cardiovascular, renal, respiratory and metabolic disorders. Failure to respond now to these changes can be very costly in terms of disease, health-care expenditure and lost productivity alongside ecological imbalance and environmental degradation. Consequently, existing action plans, policies and measures might become insufficient at higher levels of risks or in the face of more frequent and intense events.

Variations in meteorological variables are likely to have severe adverse effects on health. According to WHO global estimates, 210 million people have COPD [4]. In Europe, temperature increase will contribute to the burden of disease and premature deaths, particularly in population subgroups such as the elderly and patients with COPD [5]. Many COPD sufferers have their symptoms deteriorate during colder weather; this often leads to an increase in hospital admissions and capacity shortages [6].

In the UK, climate change might increase COPD, asthma, and tick-infected diseases (e.g. Lyme disease infected ticks during spring and autumn) and some food-borne diseases during the warmer summers [7]. Furthermore the highly significant negative association between temperature and death rates from various causes like chronic bronchitis, pneumonia, ischaemic heart disease and cerebrovascular disease indicates a change in the future disease prevalence [8]. The incidence of Hip Fracture varies seasonally [9] and adolescent's emotional distress level is found to be highly significant due to drought [10]. Most of the direct impacts were mainly based on the increase of temperature in the form of heat related mortality and heat related diseases like dehydration, heat-stroke, and hyperthermia [11], [12].

In this paper, we explore the association between incidence rates and monthly maximum, minimum, mean temperature and monthly total rain by using data for April 1997 to March 2004 for a particular region

(England North). We use a statistical model (zero-inflated Poisson regression model) to measure the significance of meteorological variables on COPD admission counts [13, 14].

Poisson regression models provide a standard framework for the analysis of the count data. In practice, however, count data often have a higher incidence of zero counts than is expected for the Poisson distribution. This phenomenon, also called, over dispersion, occurs because a single Poisson Parameter (μ) is often insufficient to describe the population. One approach to analyzing count data with many zeros is to use Zero-inflated Poisson distribution [13, 14].

Our preliminary findings may act as evidence to support the hypothesis, that there is an association between weather changes and COPD incidence rates. Subsequently, it might be possible to forecast the future demand for care services based on specific diseases and regional variations in meteorological variables.

The paper is organized in the following sections: data description is in section 2; a description of the calculation of incidence rate is given in section 3. Section 4 illustrates the trends in COPD incidence rates, which is linked to meteorological factors based on a single month over a number of years. Sections 5 describe the fitted model for COPD admission counts, and finally our paper is concluded in section 6.

2. Data

In this study, two datasets are used, namely the national Hospital Episodes Statistics[1] (HES) dataset and the observational data from the Met Office[2]. We also collected the mid-year population for a number of years from the Office for National Statistics, UK[3].

HES is a data warehouse containing all the admission statistics of personal, medical and administrative details of all patients admitted to, and treated in, NHS hospitals in England. It also contains details of all NHS outpatient appointments. There are approximately 12 million records for each financial year (in UK, a financial year is from 1 April to 31 March to the following year). The dataset includes all the consultant episodes of a patient during their stay in hospital by using specific variables under different categories (e.g. admissions, augmented/critical care period, clinical, geographical). Since our aim is to study patients with COPD (ICD-10 codes J40-J44) as their first diagnosis, we select all the COPD patients for the financial year of 1997 to 2003 for region England North.

The Met Office is the UK's national weather service, and deals with weather predictions, forecast, climate change and weather science research. The Met Office provides services to various sectors in UK and the world (e.g. industry, transport, government, health, and defense). For this study, monthly observational data for temperature (maximum, minimum and mean) and rain have been collected for the study period and region (April 1997 to March 2004).

Data was imported to MySQL version 5.0[4] and necessary steps were taken to build a longitudinal dataset (1997-2004), where meteorological data was linked for each inpatient admissions based on matching region and time (e.g. month and year).

We used the admission date for selecting the time of COPD admission and government office region of residence for the location of the patient (e.g. England North East, England North West).

3. Incidence Rate of COPD

The incidence rate is the number of new cases of disease during a period of time divided by person-time at risk of the observation period.

$$\boldsymbol{Incidence\ rate} = \frac{Number\ of\ events\ occuring\ during\ observation}{Person - time\ of\ follow - up}$$

Total person time at risk for a given subcategory is calculated by adding the person time at risk counted in that subcategory for each subject in the analysis. For our calculation, we used the person-days of follow up for calculating the COPD incidence rate. Our follow up period starts at the beginning of each month (from April 1997 to March 2004) until the end of month, e.g. 30 days on average. For instance, for January 1998 the Incidence rate is:

$$\boldsymbol{Incidence\ rate}_{Jan,1998} = \frac{Number\ of\ COPD\ admissions\ in\ January\ 1998}{Total\ Person - days\ of\ follow - up\ for\ January\ 1998}$$

If we have 3 cases of COPD having diagnosed at 10th , 15th and 20th day of the month of January, then the total person-days of follow-up would be 45

[1] www.hesonline.nhs.uk
[2] www.metoffice.gov.uk
[3] www.statistics.gov.uk
[4] www.mysql.com

(10+15+20) and the Incidence rate would be 6.67 per 100 person-days.

We multiplied the incidence rate by 100 and expressed it as per 100 person-days of follow up. The incidence rates for various months and years are illustrated in Table 1. It is evident from the Table 1 that most of the time the incidence rate increases in January but decreases markedly during March and December, which is also shown by the big fluctuation during this time in Figure 1. We calculate the percentage of COPD admissions for each month by using mid-year population[5] for respective years. Interesting to see that, although the percentage of COPD admissions for December is high for all the years, we have small values of COPD incidence rate. This could be due to the fact that during December, most COPD patients are admitted during the 3rd or 4th week of the month and thereby getting high value of total person-days in the numerator of COPD incidence rate.

Table 1. Incidence rates for England North

Year	Month	Incidence rate (100 person-days)	COPD Admissions (%)
	January	7.18	0.025
	February	7.41	0.022
	March	6.04	0.031
	April	7.19	0.029
	May	6.65	0.023
	June	6.97	0.021
	July	6.53	0.022
	August	6.91	0.021
1998	September	6.75	0.019
	October	6.60	0.024
	November	6.72	0.023
	December	5.65	0.047
	January	8.10	0.056
	February	7.40	0.032
	March	6.58	0.032
	April	6.71	0.034
	May	6.87	0.032
	June	6.98	0.033
	July	6.50	0.031
	August	6.81	0.029
2000	September	6.63	0.028
	October	6.55	0.035
	November	6.67	0.034
	December	6.23	0.042
	January	7.12	0.048
	February	7.71	0.039
	March	6.84	0.039
	April	6.96	0.038
	May	6.85	0.037
	June	6.98	0.033
	July	6.76	0.031
2001	August	6.96	0.030
	September	6.52	0.031
	October	6.77	0.037
	November	6.63	0.038
	December	6.05	0.043
	January	6.97	0.053
	February	7.58	0.041
	March	6.85	0.038
	April	7.06	0.037
	May	6.68	0.037
	June	7.11	0.033
	July	6.60	0.034
	August	6.89	0.032
2002	September	6.70	0.031
	October	6.68	0.037
	November	6.86	0.037
	December	6.15	0.044
	January	6.84	0.046
	February	7.21	0.037
	March	6.58	0.041
	April	6.83	0.044
	May	6.54	0.039
	June	7.18	0.038
	July	6.68	0.037
	August	6.88	0.033
2003	September	6.91	0.037
	October	6.09	0.047
	November	7.18	0.053
	December	6.12	0.062

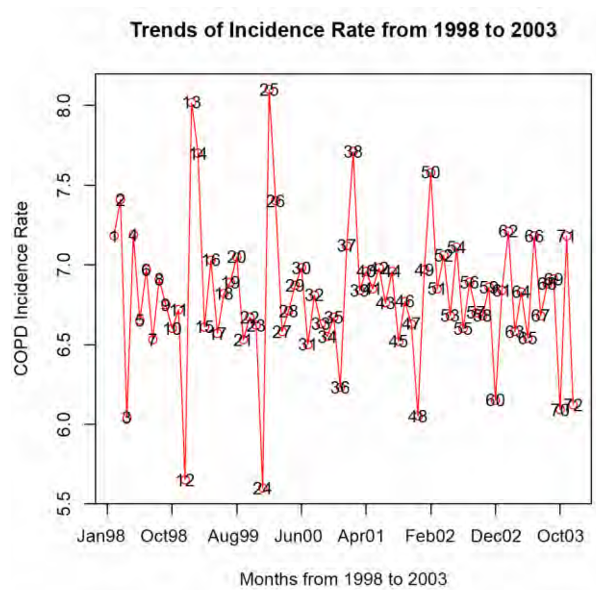

Figure 1. Time Series of COPD incidence rates for England North

---

5 www.ons.gov.uk

Table 2. Mean monthly incidence rates

Months	Incidence rate
January	7.15
February	7.51
March	6.75
April	6.96
May	6.68
June	7.0
July	6.65
August	6.93
September	6.65
October	6.58
November	6.81
December	5.97

4. Exploring COPD Incidence rates with Meteorological variables

Using the data for a single region (England North), we examined the association between meteorological variables (min., max. and mean temperature and total rainfall) and COPD incidence rates for each month for the period 1997 to 2004. The corresponding correlation matrices and the corresponding P-Values for the significance tests of the correlation results are presented in Table 3 to Table 6.

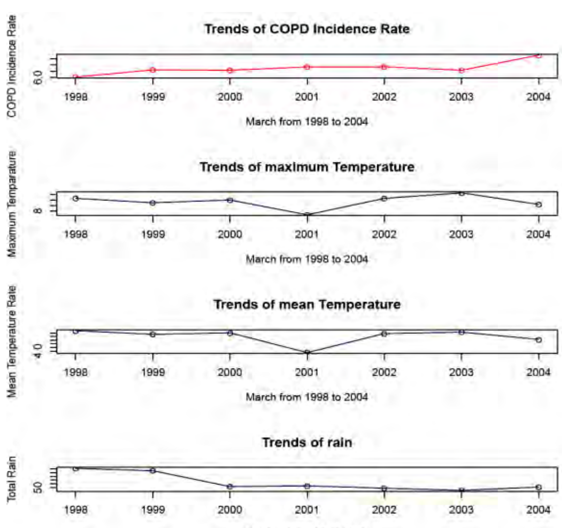

Figure 2. Trends of COPD incidence rate, maximum temperature, mean temperature and total rain for March

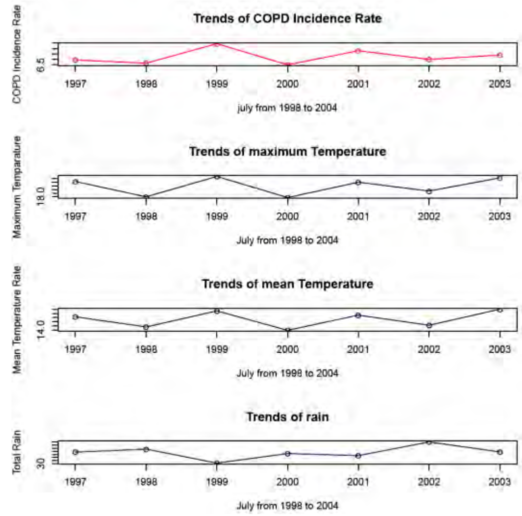

Figure 3. Trends of COPD incidence rate, maximum temperature, mean temperature and total rain for July

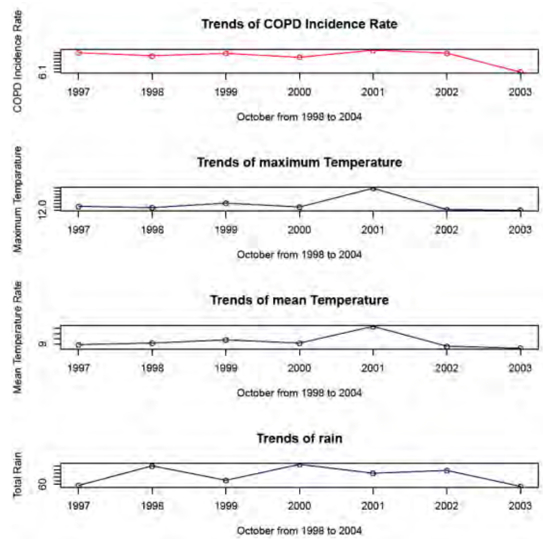

Figure 4. Trends of COPD incidence rate, maximum temperature, mean temperature and total rain for October

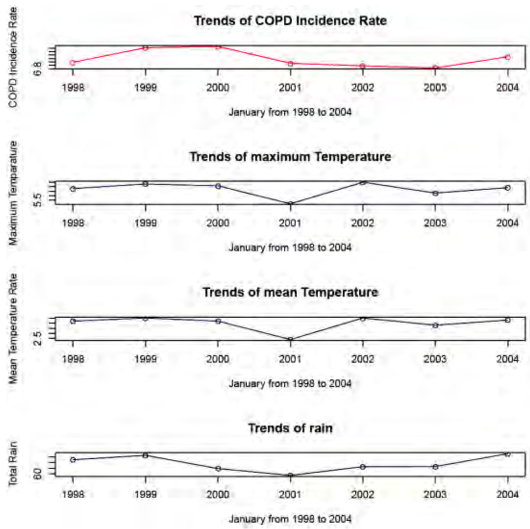

Figure 5. Trends of COPD incidence rate, maximum temperature, mean temperature and total rain for January

Table 3. Correlation matrix for March

	COPD Inc rate	Max Tem	Min Tem	Mean Tem	Rain
COPD Inc rate	1.00	-0.35	-0.39	-0.42	-0.54
Max Tem	-0.35	1.00	0.64	0.93	-0.01
Min Tem	-0.39	0.64	1.00	0.88	0.57
Mean Tem	-0.42	0.93	0.88	1.00	0.27
Rain	-0.54	-0.01	0.57	0.27	1.00
P-value		0.44	0.38	0.35	0.21

Table 4. Correlation matrix for July

	COPD Inc rate	Max Tem	Min Tem	Mean Tem	Rain
COPD Inc rate	1.00	0.82	0.68	0.80	-0.69
Max Tem	0.82	1.00	0.86	0.98	-0.54
Min Tem	0.68	0.86	1.00	0.95	-0.46
Mean Tem	0.80	0.98	0.94	1.00	-0.54
Rain	-0.69	-0.54	-0.46	-0.54	1.00
P-Value		0.02	0.09	0.03	0.08

In Figure 2, (March) the trends of incidence rates for different years are negatively related to the meteorological variables. For example, the mean temperature is increasing from 2001 to 2002 and for the same period the incidence rate is slightly decreasing. Inversely, for the period 2003 to 2004 the mean temperature is decreasing but the incidence rates is increasing. This is also true for total rainfall. For the same period the weak negative relationship can be observed from the correlation matrix in Table 3, where the correlation between incidence rate and mean temperature is -0.42 and incidence rate and rainfall is -0.54.

Table 5. Correlation matrix for October

	COPD Inc rate	Max Tem	Min Tem	Mean Tem	Rain
COPD Inc rate	1.00	0.52	0.59	0.56	0.38
Max Tem	0.52	1.00	0.97	0.99	0.07
Min Tem	0.59	0.97	1.00	0.99	0.29
Mean Tem	0.56	0.99	0.99	1.00	0.2
Rain	0.38	0.07	0.29	0.2	1.00
P-value		0.23	0.16	0.19	0.41

Table 6. Correlation matrix for January

	COPD Inc rate	Max Tem	Min Tem	Mean Tem	Rain
COPD Inc rate	1.00	0.40	0.31	0.35	0.34
Max Tem	0.40	1.00	0.91	0.98	0.61
Min Tem	0.31	0.92	1.00	0.97	0.79
Mean Tem	0.35	0.98	0.97	1.00	0.72
Rain	0.34	0.61	0.79	0.72	1.00
P-value		0.37	0.51	0.45	0.45

In July (Figure 3 & Table 4), the temperature is moderately positively correlated and the rain is moderately negatively correlated to COPD incidence rates.

During October (autumn season), we notice a positive trend between the incidence rates and meteorological variables (Figure 4). Moderate positive correlation is observed for temperature, whereas weak correlation for rainfall (Table 5). From the p-values of all the tables it is evident that, except the correlation of COPD incidence rate with maximum temperature and mean temperature for the month of July, all the correlations are not statistically significant (Table 3 to Table 6).

5. Model Fitting

A random sample of 5000 inpatient COPD admissions was selected from the HES dataset, (England North only) for the year 2003-04. We calculate all the counts of the number COPD admissions for them and the corresponding meteorological variables as illustrated in Table 7. The mean and variance of COPD admission counts are 0.023 and 0.041, respectively.

Table 7. Sample of COPD admission counts for England North 2003-04

HES ID	COPD counts	Max Tem	Min Tem	Mean Tem	Rain
100000…	0	7.4	1.7	4.6	66.1
10.	0	12.0	4.5	8.2	48.2
1000…	1	12.0	4.5	8.2	48.2
1001…	1	6.9	1.7	4.3	130.4
10000…	0	9.2	2.1	5.6	51.3
…….	…….	…….	…….	…….	…….
…….	…….	…….	…….	…….	…….

From the histogram of COPD admission counts (Figure 6), we can see that due to huge proportion of zeros, Zero-Inflated Poisson regression model would give a better fit of the data.

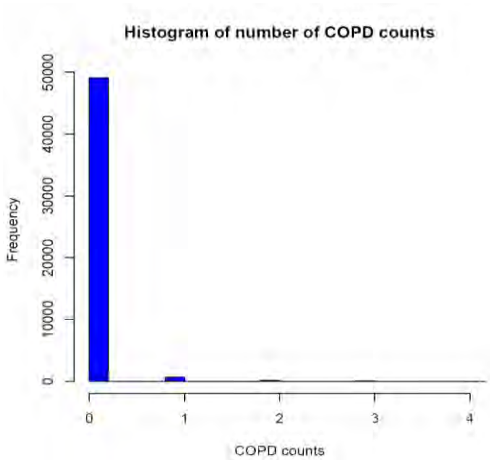

Figure 6. Histogram of COPD counts

Therefore we fit the following zero-inflated Poisson regression model:

$$P(Y = y_i|x_i, z_i) = \varphi_i + (1-\varphi_i)\, f(\mu_i, \alpha; 0), \quad y_i = 0$$
$$= (1-\varphi_i)\, f(\mu_i, \alpha; 0), \qquad y_i > 0$$

[13]. Where $f(\mu_i, \alpha; 0), y_i = 0,1,2, \ldots$is the Generalized Poisson Regression model representing the COPD admissions counts and $0 < \varphi_i < 1$. $z_i = (1, z_{i2}, \ldots z_{im})$ is the i-th row of the covariate matrix $\boldsymbol{Z}$ constitute here with the monthly maximum temperature, monthly minimum temperature, monthly mean temperature and monthly total rain. Here the two parameters of the model φ_i and μ_i depend on the covariates x_i and z_i respectively where φ_i represents the probability of the state of the structural zeros that is, the amount of the structural zeros above and beyond the sampling zeros expected by the Poisson distribution. Again, α and μ_i are the parameters of the generalized Poisson distribution stands for dispersion parameter and the mean of y_i.

Table 8. Count model coefficients (Poisson with log link)

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	0.296	0.575	0.515	0.607
Max. Tem	0.719	1.08	0.663	0.507
Min. Tem	0.859	1.079	0.797	0.426
Mean Tem	-1.574	2.156	-0.73	0.465
Rain	-0.005	0.003	-1.71	0.087

Zero-inflated Poisson regression model predicting the number of COPD counts from maximum temperature, minimum temperature, mean temperature and total rain was statistically significant (from likelihood ratio test, chi-squared= 47.602, p<0.01). None of the predictor variables are found to be statistically significant for COPD admissions counts. We perform the Vuong test (test statistics = -10.22 and p-value < 0.0000), which suggests that the zero-inflated model has a significant improvement over Poisson model.

6. Conclusion

This paper focused on the interrelationship between incidence rates of COPD admission and some selected meteorological variables by comparing their trends with time and calculating the corresponding correlation matrices. Furthermore, we also provided the model fitting to test the significance of meteorological variables to COPD admission counts. The result shows that COPD has a high incidence rate during winter, which also varies a lot during that season. Again, although the trends of COPD incidence have a moderately negative correlation with temperature and rain throughout the year, the outcomes of the model fitting (zero-inflated Poisson regression model) reveals the lack of statistically significant relationship of meteorological variables on COPD admissions counts (2003-04, England North). Interesting to note that December had the highest number of COPD admissions on average but the incidence rate is the lowest compared to other months.

We have only considered COPD based on a single region, where the analysis could easily be replicated to other regions and disease categories which are thought to be affected due to weather variations. This could show the case for regional variations in such situations, which could then be used to revise healthcare policies depending on changing climate.

7. References

[1] Betts R., *Science must end climate confusion*, online: http://news.bbc.co.uk/1/hi/8451756.stm, January 11, 2010, Accessed February 24, 2010.

[2] WHO Europe, *Protecting Health in Europe from Climate Change*, World Health Organization report, 2008.

[3] WHO Europe, *Taking action to protect health in Europe from Climate Change*, Fact Sheet, http://www.euro.who.int/mediacentre/FactSheets/FactSheets, 2008.

[4] WHO, *Chronic obstructive pulmonary disease (COPD) Fact sheets*, http://www.who.int/mediacentre/factsheets/fs315/en/index.html, 2009.

[5] ERS, *Climate Change and respiratory disease*, European Respiratory Society report, 2010.

[6] P. Marno, C. Bryden, W. Bird, and H.A. Watkin, "How different measures of cold weather affect chronic obstructive pulmonary disease (COPD) Hospital Admission In London", *European Respiratory Review*, 2006, 15:101, pp.185-186.

[7] DH and HPA, *Health Effects of Climate Change in UK 2008: An update of the Department of Health report 2001/2002*, Department of Health and Health Protection Agency report, 2008.

[8] I.H. Langford, G. Bentham, "The Potential effects of climate change on winter mortality in England and Wales", *International Journal of Biometeorol*, 1995, 38, pp. 141-456.

[9] Tenías J. M., Estarlich M., Fuentes-Leonarte V., Iñiguez C., Ballester F., "Short-term relationship between meteorological variables and hip fractures: An analysis carried out in a health area of the Autonomous Region of Valencia, Spain (1996–2005)", *Bone*, 2009, 45, pp. 794–798.

[10] J. G. Dean, H. J. Stain, "Mental health impact for adolescents living with prolonged drought", *Australian. Journal of Rural Health*, 2010, 18, pp. 32–37.

[11] A. Fouillet, G. Rey, F. Laurent, et al, "Excess mortality related to the August 2003 heat wave in France", *International Arch Environ Occup Health*, 2006, 80, pp. 16–24.

[12] C. Stollberger, W. Lutz, J. Finsterer, "Heat-related side-effects of neurological and non-neurological medication may increase heat-wave fatalities", *European Journal of Neurology*, 2009, 16, pp. 879–882.

[13] F. Famoye, K.P. Singh, "Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data". *Journal of Data Science*, 2006, 4, pp. 117-130.

[14] M. Ridout, J. Hinde, C. G. B. Demétrio, "A Score Test for Testing a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alternatives". *BIOMETRICS*, 2001, 57, 219-223.